**Other Reasons for Adopting Cluster Randomization** To enhance subject compliance Community-level effectiveness trials Administrative convenience To obtain cooperation of investigators Ethical considerations

#### Impact on Design and Analysis



Inferences in CRTs are frequently intended to apply at the individual level, while randomization is applied at the cluster level.



- Problem with individual level analysis: lack of independence among members in a cluster (clustering effect).
- Application of standard sample size formulas will lead to underpowered studies.



Application of standard statistical methods will tend to bias p-values downward risking a spurious claim of statistical significance.

# How does clustering operate to produce these effects?

- Non-independence: people in clusters tend to be more similar to each other that to persons in a different cluster (ages, sex ratios, ethnicities, income)
- Environmental factors are more similar within clusters that across clusters (treatment procedures and success rates within a clinic, differences in temperature between nurseries related to infection rates).
- Interaction between individuals within clusters, as a result, may respond similarly to interventions

## Measuring the degree of Between Cluster Variability

- Between cluster variability and within cluster similarity are two ways of viewing the same nonindependence phenomenon.
- Coefficient of Variation (k):
  a measure of the between cluster variability
  The coefficient of variation k, can be defined as

 $k = \sigma_B / \mu$ 

where

 $σ_B$  = between-cluster variance among c clusters on  $μ_j$ subscript j goes from 1 to c μ = true value of parameter of interest Measuring the degree of the Clustering Similarity

Intra-class correlation coefficient "rho" (ρ):
 a measure of the degree of similarity among responses within a cluster

The overall response variance  $\sigma^2$  may be expressed as the sum of two components, i.e.,

 $\sigma^2 = \sigma^2_A + \sigma^2_W,$ 

where

 $\sigma^{2}_{A}$  = between-cluster component of variance  $\sigma^{2}_{W}$  = within-cluster component of variance

then

 $\rho = \sigma_A^2 / (\sigma_{A+}^2 \sigma_W^2)$ 

# Design Effect or Variance Inflation Factor (VIF)

For simple sample size determination, "design effect" is defined as:

#### $DE = 1 + (m - 1) \rho$

(where ρ is the intra-class correlation and m is the cluster size number of subjects within a cluster).

Design Effect/VIF gives a measure of how much the sample size in each group have to be increased to achieve the same statistical power as would be obtained by individual level randomization.

## Problem

In order to estimate k or rho you have to have fairly extensive prior knowledge (read data) on the outcomes you want to study and how they distribute themselves in the clusters of population members you intend to randomize.

… as well as make some sophisticated guesses about likely intervention responses and heterogeneity.

#### Answers

Find ICC's or k's from studies in the same of similar settings.

There is a push in the literature for investigators to publish ICC's from cluster randomized trials

Perform pilot studies to estimate ICC's & k's.